Journal of Geochemical Exploration 141 (2014) 61–70

Contents lists available at ScienceDirect

Journal of Geochemical Exploration

journal homepage: www.elsevier.com/locate/jgeoexp

Linking compositional data analysis with thermodynamic geochemical modeling: Oilfield brines from the Permian Basin, USA

Mark A. Engle [a,b,*], Madalyn S. Blondes [a]

[a] U.S. Geological Survey, Reston, VA, USA
[b] Dept. of Geological Sciences, University of Texas at El Paso, El Paso, TX, USA

ARTICLE INFO

Article history:
Received 17 July 2013
Received in revised form 17 February 2014
Accepted 20 February 2014
Available online 3 March 2014

Keywords:
Permian Basin
Produced water
Oilfield brine
Compositional data analysis
Geochemical modeling

ABSTRACT

Compositional data analysis (CoDa) of sedimentary basin brines can help reveal the hydrogeochemical history of hydrocarbon producing formations. CoDa uses log-ratio transformations to convert compositional data into Euclidean space, a necessary assumption for many mathematical models and basic data interpretation. This paper demonstrates the novel approach of pairing CoDa with thermodynamic geochemical models for bivariate and multivariate subcompositions for a suite of brines within three producing formations from the Permian Basin, USA. Subcompositional analysis of basin brines using [Na, Cl, H_2O] and [Ca, SO_4, H_2O] with corresponding equilibration models, plotted using isometric log-ratio (ilr) transformations, shows reaction with anhydrite and halite at depth. The [Na, Cl, Br] subcomposition often can discriminate between seawater evaporation and halite dissolution; the sparse Br data suggest that both processes occur. Principal component analysis (PCA) of centered log-ratio (clr) data is useful for visualizing combinations of processes that are not readily apparent in the two-solute models. PCA of the Permian Basin brines reveals distinct groups of data, including samples likely affected by water flooding, and others influenced by a combination of Late Permian seawater evaporation or halite dissolution coupled with Ca + Mg ion exchange for Na on clays. These relationships support a hydrogeologic model in which meteoric inflow occurs in a heterogeneous manner much more complicated than the previously suggested concept of west-to-east piston style flow.

Published by Elsevier B.V.

Introduction

Collaborations between mathematicians and geochemists have provided new tools for exploring and analyzing geochemical data that appropriately account for their compositional nature (Aitchison, 1986; Buccianti and Pawlowsky-Glahn, 2005; Egozcue et al., 2003; Filzmoser et al., 2009; Templ et al., 2008). This approach, compositional data analysis (CoDa), uses a family of log-ratio transformations (Aitchison, 1986; Egozcue et al., 2003) to convert the original compositional data, which sit in the sample space of the simplex, into coordinates that follow the rules of Euclidean geometry in real space. Following these rules is a necessity for many statistical and mathematical models, including correlations and relationships within bivariate subcompositions, because this is the geometry upon which these methods rely (Egozcue and Pawlowsky-Glahn, 2005; Egozcue et al., 2003; Filzmoser et al., 2010). Depending on the particular transformation employed, a dataset of D parameters, or parts, is recast into D or D-1 number of new variables that can be examined and interpreted directly. However, patterns and trends in the transformed data require a different style of interpretation than what is traditionally taught in the geosciences, leading to potential confusion. Several authors have tried to remedy this situation by providing intuitive examples (Buccianti, 2011; Pawlowsky-Glahn and Egozcue, 2011) or linking the structure of the log-ratio transformations to real-world concepts such as the stoichiometric relationships between elements in mineral lattices (Grunsky et al., 2008). We propose that in addition to typical CoDa approaches for bivariate and multivariate subcompositions (Aitchison, 1986; Buccianti, 2011; Filzmoser et al., 2009, 2010), mapping of similarly transformed geochemical reaction pathways and mineral saturation data into scatterplots and principal component (PC) biplots can provide more robust and meaningful interpretations while still utilizing proper methods for the treatment of compositional data.

The similarities between the structure of log-ratios and equations for the Law of Mass Action have previously been noted (Buccianti et al., 2005) but efforts to link thermodynamics and CoDa are limited. To demonstrate the utility in linking the two, this investigation applies geochemical modeling pathways and saturation indices to the CoDa of a relatively simple system: major and minor ions found in brines produced with oil and gas resources from the Permian Basin, Texas

---

* Corresponding author at: Dept. of Geological Sciences, University of Texas at El Paso, El Paso, TX, USA. Tel.: +1 915 747 5503; fax: +1 915 747 5073.
E-mail address: engle@usgs.gov (M.A. Engle).

http://dx.doi.org/10.1016/j.gexplo.2014.02.025
0375-6742/Published by Elsevier B.V.

and New Mexico, USA. Despite the immiscibility between water and oil, many hydrocarbon reservoirs, particularly mature oil plays, co-produce large quantities of water (Clark and Veil, 2009). Sources of this water include formation brine that naturally occurs in sedimentary basins, water that condenses out of the gas phase during production, and water injected in the reservoirs during drilling, stimulation, or water flooding (enhanced oil recovery). Studying the chemistry of produced waters from hydrocarbon reservoirs can produce insights into the source of water, fluid flow and reservoir connectivity, thermal maturation, and diagenesis (Barnaby et al., 2004; Bein and Dutton, 1993; Carpenter, 1978; Dutton, 1987; Engle and Rowan, 2013, in press; Stueber et al., 1998; Walter et al., 1990). This current effort is intended to demonstrate how modeling results can be used to ease data interpretation of log-ratio transformed produced water compositional data. This work focuses on both bivariate and multivariate subcompositions.

Compositional data and log-ratio transformations

Compositional data are defined as multivariate observations with positive real values in which the components, or parts, are relative in nature and subsets of a whole, κ. They are constrained to a hyperplane in real space known as the simplex, S^D, which is defined for a D-part composition $\mathbf{x} = (x_1, \ldots, x_D)$ as

$$S^D = \left\{ \mathbf{x} = (x_1, \ldots, x_D), x_i > 0, i = 1, \ldots, D, \sum_{i=1}^{D} x_i = \kappa \right\}. \quad (1)$$

Essentially, nearly all geochemical data follow this definition and thus are considered compositional, regardless of their units (e.g., mol kg^{-1}, mg L^{-1}, etc.) (Aitchison, 1986; Buccianti and Pawlowsky-Glahn, 2005; Egozcue et al., 2003; Engle and Rowan, 2013, in press). Ionic and elemental data from water samples fall under this scenario. An important property of CoDa is subcompositional coherence. This property assures us that relationships between parts in one subcomposition are identical to those between the same parts in the full composition or any other subcomposition. This same coherence is not guaranteed in conventional data analysis techniques (Egozcue, 2009; Geboy et al., 2013). To properly account for the constraints on compositional data and to fulfill properties of CoDa including subcompositional coherence, the results are typically transformed using either the additive log-ratio (alr), centered log-ratio (clr), or isometric log-ratio (ilr) transformations.

In the case of the alr transformations, one part (the jth part) is selected as the divisor which is applied to the other D-1 constituents of $\mathbf{x}$:

$$\mathrm{alr}(\mathbf{x}) = \left(\ln\frac{x_1}{x_j}, \ldots, \ln\frac{x_{j-1}}{x_j}, \ln\frac{x_{j+1}}{x_j}, \ldots, \ln\frac{x_D}{x_j} \right). \quad (2)$$

For water geochemistry data, one simple choice for the divisor is the mass of water. In dilute solutions where the sample density ~1 kg L^{-1}, one can approximate this transformation simply by taking the natural log of concentration data in mass per volume or mass per mass units. In the case of highly saline fluids, such as those investigated here, water content is determined by taking the difference between the solution density and the total dissolved solids content. The alr transformation has been used for regression, hypothesis testing, and plotting of three-part subcompositions (Aitchison, 1986; Thomas and Aitchison, 2006), but does suffer from distortion given that this transformation produces coordinates with respect to an oblique basis (Mateu-Figueras et al., 2011). However, because of the transformation's mathematical simplicity and its visual similarity to commonly used ratio/ratio plots, it is often used to make preliminary interpretations.

The structure of the clr transformation has some similarities to the alr transformation although each of the D-parts of $\mathbf{x}$ is divided by the geometric mean of all of the parts in the dataset and a total of D clr transformed variables are produced (rather than D-1 as with the alr transformation):

$$\mathrm{clr}(\mathbf{x}) = (y_i, \ldots, y_D) = \left(\ln\frac{x_i}{\sqrt[D]{\prod_{i=1}^{D} x_i}}, \ldots, \ln\frac{x_D}{\sqrt[D]{\prod_{i=1}^{D} x_i}} \right). \quad (3)$$

Although the clr transformation provides a 1-for-1 correspondence for each part, the resulting sample covariance matrix is singular and relationships between the constituents vary depending on which parts are included in the denominator. As such, this transformation is less useful for co-plotting of variables (see Drew et al., 2008) and is more commonly used for covariance-based PC analysis.

The ilr transformation converts compositional data to a Euclidean space like alr and clr, preserves distances like clr, but uniquely has an associated orthonormal basis and preserves angles between data (Egozcue et al., 2003; Filzmoser et al., 2010). While many examples have been provided showing how to interpret ilr transformed data (cf. Buccianti, 2011; Pawlowsky-Glahn and Egozcue, 2011), we recognize the hesitancy of some scientists to use the ilr transformation given its apparent mathematical complexity. However, with further clarification and care to define coordinates in a geologically intuitive manner, ilr transformed data can be more easily interpreted. In geochemical systems, knowledge of stoichiometric relationships can be leveraged in the creation of ilr balances (Grunsky et al., 2008). There are multiple methods for defining ilr coordinates but one of the more popular is the creation of non-overlapping groups of parts via a sequential binary partition (Egozcue and Pawlowsky-Glahn, 2005). For a simple 3-part subcomposition, $\mathbf{x}_3$, in which water is one of the constituents of interest (x_3), a conversion to two new ilr balances, z_1 and z_2, can be calculated as:

$$\mathrm{ilr}(\mathbf{x}_3) = (z_1, z_2) = \left(\frac{1}{\sqrt{2}} \ln\frac{x_1}{x_2}, \frac{\sqrt{2}}{\sqrt{3}} \ln\frac{\sqrt{x_1 x_2}}{\mathrm{H_2O}} \right). \quad (4)$$

Using this approach, z_1 represents the ratio in the first two parts of the subcomposition and z_2 is a measure of the proportion of the first two parts relative to the water content of the sample. Note that z_2 is similar in structure to the equation for calculating the ion activity product (IAP) of a simple salt, for example anhydrite:

$$\mathrm{IAP} = \left[\mathrm{Ca}^{2+}\right]\left[\mathrm{SO}_4^{2-}\right] = \frac{mol_{\mathrm{Ca}^{2+}}}{v_{\mathrm{H_2O}}} \frac{mol_{\mathrm{SO}_4^{2-}}}{v_{\mathrm{H_2O}}}, \quad (5)$$

where mol is the number of moles of each ion and v is the volume of pure water in solution. Pulling out a constant (c) to account for water density and to convert from moles to mass (m) of the ions and taking the square root of both sides produces an equation nearly identical to z_2:

$$\sqrt{\mathrm{IAP}} = c\frac{\sqrt{m_{\mathrm{Ca}^{2+}} m_{\mathrm{SO}_4^{2-}}}}{m_{\mathrm{H_2O}}}. \quad (6)$$

Thus comparison of z_2 in ilr transformed data for the subcomposition [Ca, SO_4, H_2O] with the transformed equilibrium constant for anhydrite provides insight into the possible role of evaporite minerals controlling water chemistry.

Geology and brines of the Permian Basin

The Permian Basin of west Texas and southeastern New Mexico (Fig. 1) is an important source of hydrocarbons, producing more than 29×10^9 barrels (4.6×10^9 m^3) of oil and 75×10^{12} ft^3 (2.1×10^{12} m^3) of natural gas over the last ~85 years (Railroad Commission of Texas, 2013). The 300,000 km^2 geologic basin consists of several major structural features including the Northwest Shelf, Delaware Basin, Central Basin Platform, Midland Basin, and Eastern Shelf